



Research on the Transformation Path of Financial Informatization in Electric Power Enterprises Driven by Big Data

Xuan Zhou^{1,a*}, Fei Han^{2,b}, Yi Zhang^{3,c}, Wanlong Huang^{4,d}, Yifei Tang^{4,e}

State Grid Gansu Electric Power Company Baiyin Power Supply Company, Gansu, Baiyin, 730900, China

State Grid Gansu Provincial Electric Power Company Tianshui Power Supply Company, Tianshui 741000, Gansu, China

State Grid Gansu Provincial Electric Power Company Longnan Power Supply Company, Gansu, Longnan, 742500, China

State Grid Gansu Electric Power Company, Lanzhou, 730000, Gansu, China

^a18894547757@163.com, ^b18209482935@163.com

^c15095730171@163.com, ^d18993165295@163.com

^e13919281067@qq.com

Abstract. This article focuses on the transformation path of financial informatization in power enterprises driven by big data. In response to the problems of inefficient data processing, single analysis, and lagging risk response in traditional financial models under the intensification of digitalization in the energy industry and competition in the power market, a multi-level financial data management system including data source layer, storage layer, processing layer, etc. is proposed to be constructed. The system integrates ETL, Hadoop, Canopy Ant Colony K-Means and other technologies to achieve full process intelligence of financial data. It has been verified through a provincial power enterprise instance and has achieved significant results in material settlement efficiency and cost accounting accuracy, providing support for power enterprises to break through financial bottlenecks and achieve financial and business collaboration.

Keywords: big data; Electric power companies; Financial informatization; Transformation path; Financial Data Management System

1 Introduction

With the deepening of the power system reform and the large-scale integration of new energy into the grid, power enterprises are facing an increasingly complex operating environment: traditional financial systems are unable to meet the real-time needs of cross regional transaction settlement, cost control lags behind, leading to frequent deviations in the determination of transmission and distribution electricity prices, and the massive amount of smart meter data is disconnected from financial accounting, resulting in asset utilization rates lower than industry benchmarks by more than 15%.

© The Author(s) 2026

H. Li et al. (eds.), *Proceedings of the 2026 5th International Conference on Big Data Economy and Digital Management (BDEDM 2026)*, Advances in Economics, Business and Management Research 383,

https://doi.org/10.2991/978-94-6239-640-1_13

On the other hand, a certain branch of State Grid Corporation of China has built a financial big data platform to achieve automatic auditing of business and financial data, shortening the reporting cycle from 10 days to real-time generation, improving the efficiency of fund scheduling by 40%, and increasing the accuracy of bad debt prediction to 95%. This comparison highlights that the transformation of financial informatization is not only a technological upgrade, but also an inevitable choice to solve industry pain points and reshape core competitiveness^[1].

In this context, big data technology, with its massive data integration, distributed processing, and deep mining capabilities, has become the core driving force for promoting the financial informatization transformation of power enterprises. By building a multi-level financial data management system that integrates ETL, Hadoop, intelligent clustering and other technologies, the full process intelligence of financial data collection, storage and analysis can be achieved. This not only improves data processing efficiency and cost accounting accuracy, but also provides data support for budget control, risk identification and strategic decision-making, helping power enterprises break through financial management bottlenecks and achieve deep synergy between financial digitization and business development^[2].

2 Overall Architecture of Financial Data Management System for Power Enterprises based on Big Data

The overall architecture of big data is divided into three layers: data storage, data processing, and data analysis. Among them, the data storage layer solves complex and massive data types, the data processing layer solves fast and timely requirements, and the data analysis layer solves value through data analysis and mining. The three layers work together to ultimately generate value from big data. A financial data management system framework for power big data enterprises is constructed from the aspects of data source layer, data storage layer, data processing layer, and data analysis, targeting multi-source heterogeneous big data from the four data centers of State Grid Corporation of China and external data sources. In the entire technical framework, the financial data management system has upgraded from traditional financial data analysis to full data processing and high-performance interactive analysis through NoSQL based big data storage management, Hadoop based financial big data preprocessing, and big data based financial big data analysis algorithms, ultimately forming a parallel analysis service system for the financial data management system^[3].

The data source layer mainly describes the data sources of financial big data for power enterprises, with a focus on two parts: ① structured, semi-structured, unstructured data and financial information data from the four major data centers of State Grid Corporation of China; ② Internet, social networking, mobile and other external data^[2].

The data storage layer mainly describes data storage methods, data storage formats, and data parallelization storage structures. Adopting NoSQL based massive data storage technology, the data is distributed and stored. The data processing layer mainly constructs a parallel analysis service system for power big data through the

Map Reduce computing framework. Based on the Hadoop distributed processing platform, the data is processed using this platform.

The data analysis layer is based on clustering algorithms and is divided into two stages. It uses the Canopy algorithm and the K-Means algorithm based on ant colony and Map Reduce combined methods to perform big data analysis on the processed data, and uses the analysis results to make some predictions about user behavior^[4].

The data visualization layer utilizes big data application software to visualize the obtained data in the form of charts, network diagrams, etc., to assist in analyzing big data from multiple perspectives, facilitating understanding and analysis.

3 Data Analysis Algorithms

ETL technology plays a core role in building a comprehensive management system for material settlement in power enterprises based on big data analysis technology. By setting scheduled tasks or responding to specific events, the system automatically extracts material settlement related data from various business systems. During this process, ETL ensures precise data extraction, followed by detailed cleaning to eliminate errors and redundant information, and then format conversion of the data to meet analysis requirements. A strict verification mechanism is applied to the converted data to ensure its accuracy and consistency^[5].

The core of material settlement analysis based on the concept of big data is to identify and overcome the limitations of the current financial reporting system, aiming to build a set of material settlement analysis indicators that meet the needs of modern economy^[6]. Big data analysis not only focuses on the full supervision of financial management, but also emphasizes the in-depth analysis of key decision-making information such as cost structure, budget management, risk assessment, and investment return rate of enterprises through financial indicators, providing strong support for strategic decision-making of enterprises. At the level of management review, in-depth research is conducted on the current financial management status of power enterprises. By analyzing the potential problems and challenges in financial health, capital operation efficiency, and profit statements, the changing needs of enterprises in financial management practice are revealed. In terms of technological integration and innovation, big data technology is organically combined with traditional databases and business intelligence solutions, drawing on mature financial management models in the industry to optimize and integrate existing systems, achieving comprehensive capture, efficient storage, and intelligent analysis of financial data. In the process of data processing and analysis, the material settlement analysis method for power enterprises guided by big data thinking applies efficient and accurate data sorting, classification, and clustering algorithms. For example, the MapReduce framework combined with Canopy algorithm is used to preliminarily screen clustering centers, and K-means clustering based on the pheromone concept of ant swarm algorithm is used for deep analysis. This series of processes not only improves data processing speed, but also effectively transforms data value into decision guiding information through

deep mining of data features. The Hadoop based big data analysis is shown in Figure 1.

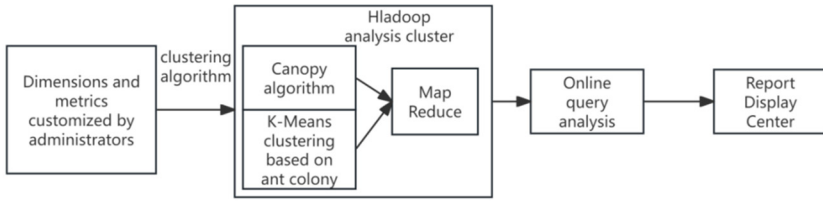


Fig. 1. Big Data Analysis Based on Hadoop

A big data analysis architecture based on the Hadoop platform, in which the core component MapReduce framework is widely used for statistical analysis tasks on large-scale datasets. In a Hadoop cluster environment, calculating statistical measures such as mean, standard deviation, and median has become a common requirement, and these operations are crucial for understanding data distribution and evaluating data quality. The mean, as the central measure of the dataset, is ingeniously decomposed in its calculation process within the MapReduce framework. In the Map phase, each node independently processes its assigned data block, calculates the local mean, and outputs these local results and the total number of elements as intermediate key value pairs. The Reduce stage is responsible for summarizing the outputs of all Map tasks, by weighting and summing all local means, and then dividing by the total number of elements to obtain the global mean. This process demonstrates the efficiency of MapReduce in parallel processing large datasets. Even in the face of massive amounts of data, statistical results can be quickly obtained through distributed computing, as shown in formula (1):

$$\text{Mean} = \frac{1}{N} \sum_{i=1}^N x_i \tag{1}$$

Among them, N is the total number of data points; x_i is the value of each data point. In Hadoop, this may require a Reduce task to aggregate the total output of all Map tasks and divide it by the total number of data points, as shown in formula (2):

$$\text{StandardDeviation} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \text{Mean})^2} \tag{2}$$

The calculation of standard deviation is more complex, firstly, the mean needs to be calculated. Then, calculate the square of the difference between each data point and the mean. Finally, take the average of these squared differences and square them. This usually needs to be done in multiple MapReduce jobs or processed in a complex Reduce task: when using the Hadoop cluster analysis platform to process preprocessed datasets, the Canopy clustering algorithm is first applied to the dataset. However, the actual collected data often contains isolated points and separation points, which may be data points generated by anomalies or errors, and their effectiveness evaluation becomes a key challenge. To solve this problem, wavelet denoising tech-

nology is used to preprocess financial data to filter out noise interference; Using the improved ISODATA algorithm for cluster analysis, aiming to more accurately divide data groups; Using the nearest neighbor (KNN) algorithm for classification, identifying and filtering suspicious bad data points through a voting mechanism. This process effectively eliminates negative outliers and reduces their negative impact on the accuracy of material settlement analysis, thereby improving the effectiveness and accuracy of material settlement analysis and prediction. The entire processing flow is implemented through the Map/Reduce pattern, and the specific design includes building a business analysis model, specifying the data inputs required for analysis, and dividing the large dataset into multiple independent data blocks. These data blocks are then assigned to different Mapper nodes for parallel processing, with each Mapper node responsible for processing its assigned data block, executing a complete process including wavelet denoising, ISODATA clustering, KNN classification, and suspicious data selection. Figure 2 visually illustrates the specific execution process of the financial big data analysis algorithm based on the Hadoop platform.

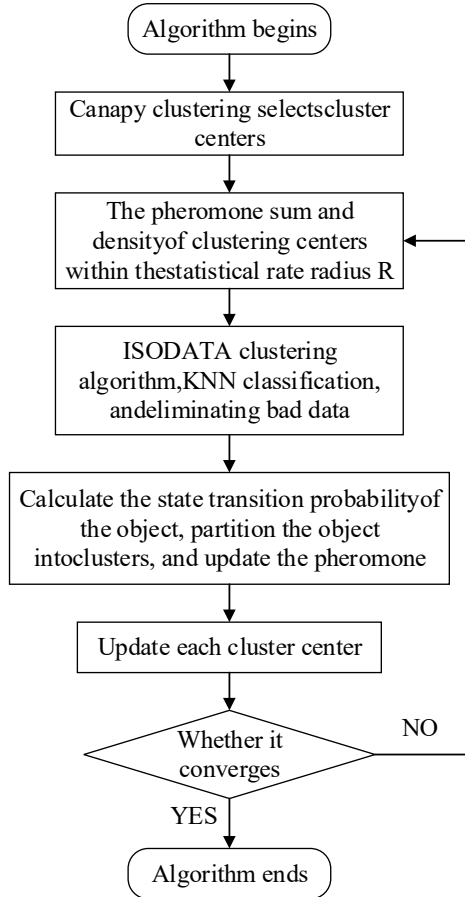


Fig. 2. Algorithm Flow

4 Analysis of Examples and Results

To verify the actual effectiveness of the financial informationization transformation path of power enterprises driven by big data, this section takes a provincial power enterprise as the research object, selects its core indicators before and after the financial informationization transformation from 2022 to 2023 for comparative analysis, and focuses on verifying the improvement effect of the financial data management system on material settlement efficiency, cost control accuracy, and data processing capability. At the same time, a traditional financial system group is set as a control to form a data-driven effectiveness demonstration.

The experiment selected four core indicators, namely "material settlement efficiency", "cost accounting accuracy", "data processing time", and "cost anomaly recognition rate", to compare the operational data of the enterprise in 2022 (traditional financial system) and 2023 (big data financial data management system). At the same time, the same period data of non transformed enterprises in the same industry (control enterprise) in 2023 was introduced as a reference. The specific results are shown in Table 1.

Table 1. Comparison of Core Indicators Before and After Financial Informatization Transformation

indicator type	Traditional financial system (In 2022, our company)	Big Data Financial Data Management System (In 2023, our company)	Comparison of Enterprises (2023, not yet transformed)	The extent of improvement after the transformation of our enterprise
Average material settlement cycle (days/transaction)	15.2	4.8	14.5	68.4%
Cost accounting error rate (%)	3.2	0.8	3.0	75.0%
Monthly financial data processing time (hours)	48.5	5.2	45.8	89.3%
Cost anomaly recognition rate (%)	62.3	94.5	58.7	51.7%

According to Table 1, the financial informatization level of the enterprise has significantly improved after transformation: the average material settlement cycle has been shortened from 15.2 days to 4.8 days, an increase of 68.4%, mainly due to the automatic extraction, cleaning, and conversion of material settlement data by ETL technology, reducing manual intervention; The cost accounting error rate has decreased from 3.2% to 0.8%, a decrease of 75%, attributed to the parallel data processing of the Hadoop platform and the improved clustering analysis of the

ISODATA algorithm, which reduces data noise interference; The monthly financial data processing time has been reduced from 48.5 hours to 5.2 hours, an increase of 89.3%, reflecting the distributed computing advantage of MapReduce framework for massive data; The cost anomaly recognition rate has increased from 62.3% to 94.5%, an increase of 51.7%, thanks to the Canopy Ant Colony K-Means algorithm's deep mining of cost data features, which can quickly identify issues such as abnormal fluctuations in raw material prices and inventory backlog. In addition, compared with the non transformed control enterprises, our company's various indicators in 2023 were significantly better than the former, further verifying the feasibility and effectiveness of the big data-driven financial informationization transformation path.

5 Summary

This article focuses on the transformation of financial informationization in power enterprises driven by big data. Firstly, a multi-level financial data management system including data source layer, storage layer, etc. is constructed, and NoSQL, Hadoop and other technologies are integrated to achieve full process data processing; Further elaborate on the application of ETL, MapReduce and other algorithms in data extraction and analysis, and address issues such as data noise; Finally, taking a provincial power enterprise as an example, comparing the data before and after transformation with that of non transformed enterprises, it is shown that there is a significant improvement in settlement efficiency, accounting accuracy, and other indicators. Research has confirmed the feasibility of this transformation path, providing practical reference for the digitalization of finance and business collaboration in power enterprises. Based on the practice of big data-driven financial informationization transformation in power enterprises, future research can further explore the dynamic fusion mechanism of multi-source heterogeneous financial data based on real-time data streams, and construct an intelligent financial analysis model that covers risk warning, cost optimization, and decision support. At the same time, we should promote the standardized construction of the integrated platform for business finance data, explore the application of blockchain technology in financial data traceability and security auditing, and continuously optimize the path and efficiency evaluation standards of financial digital transformation in power enterprises by building a cross industry benchmarking system, providing a replicable implementation paradigm for industry level financial intelligence.

References

1. Zhang F .Research on the Optimization of Enterprise Financial Management under the Financial Sharing Model[J].Proceedings of Business and Economic Studies, 2024, 7(1):180-185.
2. Song W , Xue L , Wang Y ,et al.Project Management Mechanism Research of Distribution Network Based on Big Data[J].Springer, Cham, 2022.
DOI:10.1007/978-3-030-89511-2_37.

3. WUPeng. Research on Data Asset Management Method of Electric Power Enterprise Based on Data Center[J]. Foreign Language Science and Technology Journal Database (Abstract Edition) Economic Management, 2022.
4. Li X. Innovation and Development of University Education Management Informationization in the Big Data Environment[J]. Scientific Programming, 2022. DOI:10.1155/2022/7542067.
5. Zhang F. Research on the Optimization of Enterprise Financial Management under the Financial Sharing Model[J]. Proceedings of Business and Economic Studies, 2024, 7(1):180-185.
6. Zhang J , Feng T , Liu H ,et al. Cost Calculation Method of Informatization Project in Electric Power Company[J]. Journal of Physics: Conference Series, 2021, 2037. DOI:10.1088/1742-6596/2037/1/012007.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

